Journal of Teacher Education
2017, Vol. 68(2) 213–230

DOI: 10.1177/0022487116687554
journals.sagepub.com/home/jte
SAGE

Effects of a Multimedia Professional Development Package on Inclusive Science Teachers' Vocabulary Instruction

Michael J. Kennedy[1], Wendy J. Rodgers[1], John Elwood Romig[1], John Wills Lloyd[1], and Mary T. Brownell[2]

Abstract

Vocabulary knowledge is vital for students' success in school and beyond. However, students with disabilities and others who consistently score below their peers on various measures of vocabulary knowledge have difficulties in secondary-level content area courses. Because many students with disabilities are now educated primarily in general education classrooms, their teachers report needing more professional development on instructional strategies to support this population. Using a multiple-baseline design, we tested the efficacy of a multimedia, multicomponent professional development package in which middle school science teachers in inclusive classrooms promoted science vocabulary knowledge. The professional development package improved the quality of the teachers' use of evidence-based vocabulary practices and increased the amount of time they spent explicitly teaching vocabulary in their classes.

Keywords

professional development, inservice training, students with disabilities, inclusion, teacher observation, multimedia, technology, explicit instruction, evidence-based practices

Vocabulary knowledge is critical to students' academic success. It is strongly correlated with overall academic achievement and reading comprehension (National Institute of Child Health and Human Development, 2000; Reed, Petscher, & Foorman, 2016) and contributes to the achievement gap between students with and without disabilities (Biemiller & Slonim, 2001). To illustrate, an analysis of the vocabulary subsections of the National Assessment of Educational Progress (NAEP) reading assessments shows that students with disabilities score on average lower than the bottom 25th percentile of students without disabilities (National Center for Education Statistics, 2012). This poor achievement in vocabulary has potentially large effects, which has implications for student performance in secondary content areas.

For example, in science, as vocabulary becomes increasingly complex and technical (Scruggs, Mastropieri, Berkeley, & Graetz, 2010), substantial demand is placed on reading and comprehension skills (Cervetti, Barber, Dorph, Pearson, & Goldschmidt, 2012). Although success in science requires far more than only vocabulary proficiency (Schmidt, Wang, & McKnight, 2005), there is little question of its importance (Yore, Hand, & Florence, 2004). That said, vocabulary is often challenging for students with disabilities, and performance deficits in all academic disciplines are linked to the achievement gap between these students and their peers (Mutch-Jones, Puttick, & Minner, 2012; Therrien, Taylor, Hosp, Kaldenberg, & Gorsh, 2011). In sum, few would argue new approaches are needed to support content teachers in inclusive classes when implementing evidence-based vocabulary instruction. Thus, the focus of this study is to provide professional development (PD) to middle school science teachers who have students with disabilities included in their courses; however, the vocabulary practices taught to teachers are appropriate for use with all students.

Teacher's Use of Vocabulary Instruction

Although the literature provides ample description and evidence for vocabulary practices used alone and in combination (e.g., Beck, McKeown, & Kucan, 2013), the extent to which these practices are actually present in teachers' repertoires is unclear. Numerous observational studies of general and special education teachers' literacy-related practices demonstrate many teachers are infrequent implementers of

[1]University of Virginia, Charlottesville, USA
[2]University of Florida, Gainesville, USA

Corresponding Author:
Michael J. Kennedy, University of Virginia, Bavaro Hall, Room 327, Charlottesville, VA 22904, USA.
Email: MKennedy@Virginia.edu

evidence-based practices (Ciullo, Lembke, Carlisle, Thomas, Goodwin, & Judd 2016; Klingner, Urbach, Golos, Brownell, & Menon, 2010; McKenna, Shin, & Ciullo, 2015; Swanson, 2008; Swanson, Solis, Ciullo, & McKenna, 2012). However, the observational tools used in those studies lack a high level of specificity related to vocabulary instruction.

For example, Klingner and her colleagues (2010) used the Reading Instruction in Special Education (RISE) instrument, which does not include a separate measure of vocabulary instruction, and found teachers in the study used very little direct instruction or strategy instruction. McKenna et al. (2015) and Swanson (2008) conducted syntheses of observation studies and respectively found little evidence of quality vocabulary instruction occurring within observed classrooms. To illustrate, Swanson (2008) noted, "reports of vocabulary . . . instruction were overwhelmingly missing from this corpus of studies" (p. 125). McKenna et al. (2015) found five studies that included vocabulary instruction as a measure. Of these, three reported the time spent teaching vocabulary, and two reported teachers used discussion or asking students to provide definitions as their primary vocabulary practices.

Swanson et al. (2012) gathered the most specific information about vocabulary practices used by teachers. Researchers observed more than 6,000 min of reading instruction using the Instructional Content Emphasis–Revised (ICE-R; Edmonds & Briggs, 2003) observation instrument, which broadly captures "instructional events" such as strategy instruction. They found vocabulary instruction accounted for 11% of the total observation time. Within that 11%, teachers used evidence-based vocabulary practices (e.g., morphological analysis, context analysis, mnemonic instruction, and discussion) 57% of the time devoted to vocabulary although no data are provided regarding the quality of that instruction.

Given the performance of students with disabilities on the NAEP and other more proximal measures of vocabulary performance, results from these observational studies are not surprising. Therefore, researchers and practitioners should focus on discovery, testing, and implementation of PD models that are potentially useful in helping teachers adopt and use high-quality practices. Research is particularly needed at the secondary level; in their review, McKenna et al. (2015) did not locate any vocabulary studies conducted in the secondary grades between 2000 and 2013. There are, therefore, no observation studies examining teachers' use of vocabulary evidence-based practices in inclusive secondary science classes.

PD

Research Base for PD

Research on PD to improve teacher practice is extensive, but precise prescriptions for supporting practice are scarce (Borko, 2004; Hill, Beisiegel, & Jacob, 2013; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). There is general consensus that the majority of PD teachers receive costs districts excessive amounts of money but is largely ineffective (Hill et al., 2013). Limited research shows PD for teachers can affect instructional practices and subsequently affect student outcomes (Yoon et al., 2007), but little is known about the specific features or dosage that lead to those elusive positive outcomes (Hill et al., 2013).

In recent years, experts in the field of teacher PD have called for a renewed research focus on identifying general features of PD that are effective. To illustrate, Hill et al. (2013) proposed a multistage approach to studying the features of PD in which researchers begin by studying a program at one school before expanding to multiple sites using multiple facilitators. In Stage 1, the program is piloted as an intact package. Stage 2 studies are then developed to tease apart impacts of individual features of the PD package. Stages 3 and 4 focus on experimental testing of the package under different circumstances. The current study fits in Stage 1 of Hill et al.'s proposed approach, defined as "a brief one-site pilot to ensure the feasibility of the program" in which the main question is whether the PD program is acceptable to teachers and has potential impact on their instruction and/or knowledge (p. 479). Therefore, in accordance with the recommendations for Stage 1 research, the current project studied the feasibility of a four-feature PD program called the Content Acquisition Podcast–Professional Development Package (CAP-PD; Rodgers, Kennedy, Alves, & Romig, 2017). This section reviews the research on PD as it relates to the four components of the CAP-PD package.

Feature 1: Multimedia Instruction

One area of PD research supported by empirical evidence is the use of instruction using asynchronous computerized media compared with in-person workshops or presentations of content (Blanchard, LePrevost, Tolin, & Gutierrez, 2016; Hill et al., 2013). Online or computerized PD is attractive as a PD delivery method because it can be easily individualized, allows for flexible scheduling, and is usually less expensive to implement than face-to-face PD (Dede, Ketelhut, Whitehouse, Breit, & McCloskey, 2009). Computer-based PD can be implemented in various ways; for example, it may be facilitated or independent, and single sitting or multisession.

Studies in this area (see Fisher, Schumaker, Culbertson, & Deshler, 2010; Fishman et al., 2013) indicate a pattern in the research on multimedia PD—researchers generally find essentially no differences on outcomes for students or teachers when comparing online or virtual PD delivery methods with in-person PD, even in instances where teachers indicate higher satisfaction and/or greater learning with in-person PD (Hill et al., 2013). These findings have important implications for developers of PD programs and support the current study's use of asynchronous virtual instruction on vocabulary practices.

Use of multimedia in the current study. In this study, researchers developed and implemented Content Acquisition Podcasts–for Teachers (CAP-Ts) with Embedded Modeling Videos (CAP-TVs) to bolster participants' implementation of five key evidence-based practices for vocabulary instruction (student-friendly definitions, using examples and nonexamples, breaking terms into morphological parts, highlighting semantic relationships among and between words, and having rich discussions). These practices (although not an exhaustive list) are well known to the fields of general and special education as being key for supporting student vocabulary performance (Graves, 2004; Jitendra, Edwards, Sacks, & Jacobson, 2004). Each CAP adheres to Mayer's (2009) Cognitive Theory of Multimedia Learning and is a combination of still images, occasional on-screen text, and narration (a sample CAP can be seen at https://vimeo.com/143392009). CAPs are easy to create and use and can be personalized for any purpose (see Kennedy, Alves, & Rodgers, 2015).

Since 2011, there are 14 published studies demonstrating the impact of CAP-Ts and CAP-TVs on preservice teachers' declarative, conditional, and procedural knowledge of various practices or other content in special education (see Kennedy et al., 2016). This study is an extension and replication of that body of work, but it is the first to use combinations of CAP-TVs to support teachers' implementation of vocabulary practices.

Each CAP-TV also builds a learner's declarative, procedural, and conditional knowledge as defined by Alexander, Schallert, and Hare (1991). According to their definitions, declarative knowledge is factual information, procedural knowledge is knowledge of processes or routines, and conditional knowledge is the understanding of "when and where [declarative and procedural] knowledge would be applicable" (p. 323). Building declarative, procedural, and conditional knowledge is important because teachers are unlikely to implement a new practice if they do not (a) understand its purpose and agree that it is worthwhile, (b) master the specific steps needed to plan and use the intervention or practice, (c) understand the practice well enough to know when and with whom it should be used, and (d) receive feedback on performance (Gersten, Vaughn, Deshler, & Schiller, 1997; Klingner, 2004). Although establishing declarative and conditional knowledge are steps often skipped during PD (Desimone, 2009), they are critical for changing teacher practice (Carlisle, Kelcey, Rowan, & Phelps, 2011).

The CAP-T part of each video builds declarative and conditional knowledge by teaching the specific steps of each practice and providing guidance on when they should be deployed (see the first 6 min of https://vimeo.com/143392724). The video in each CAP-TV reinforces the declarative and conditional knowledge, and also builds procedural knowledge, which supports implementation (see the second half of the above video). In sum, the CAP-TV intervention aims to support a teacher's readiness to make appropriate decisions about vocabulary instruction, from which terms to explicitly teach, to selecting practices that make sense for teaching each term, and then prepares them to implement those practices with students.

Feature 2: Modeling Videos

The use of modeling is a well-established instructional technique shown to be effective in multiple environments (Rosenshine, 2012). When students learn new actions or procedures, it is helpful to see a model of the procedure being accurately performed (Archer & Hughes, 2011). The use of video in teacher PD has been studied extensively, but almost exclusively in the context of reflecting on or analyzing practice after the implementation of certain teaching methods rather than modeling the practices in advance of their implementation (Marsh & Mitchell, 2014; Zhang, Lundeberg, Koehler, & Eberhardt, 2011). Hill et al. (2013) reviewed a subset of these studies comparing the use of videos of teachers' own versus other teachers' lessons and concluded that results of these studies indicate teachers often prefer to watch videos of their own practice, but "watching videotapes of expert teaching may provide greater benefits to knowledge and skills" (p. 484).

Another theoretical strand of video use within PD grew from Brown, Collins, and Duguid's (1989) model of cognitive apprenticeship. Cognitive apprenticeship models can occur when a learner has authentic, situated instruction models and coaching at his or her disposal (Brown et al., 1989). As an example, in this study, the CAP-PD process is an example of a cognitive apprenticeship model because teachers receive instructional materials tied explicitly to their curriculum and also have high-quality models to learn from. We address the modeling video and coaching aspects of cognitive apprenticeship in the next sections.

Use of modeling videos in the current study. There is evidence that modeling videos used in combination with direct instruction can be effective for supporting preservice and inservice teachers' practice. To illustrate, the CAP-TV model was used in three studies to bolster teachers' implementation of evidence-based practices. Ely, Kennedy, Pullen, Williams, and Hirsch (2014) and Ely, Pullen, Kennedy, and Williams (2015) used CAP-TV to teach teachers components of an evidence-based practice for elementary-level vocabulary instruction. In both studies, teachers who learned using CAP-TV improved their knowledge as measured on a researcher-created instrument, and then implemented the practice with fidelity as measured using a checklist and rating of the practice's specific steps.

Kennedy, Hirsch, Rodgers, Bruce, and Lloyd (2017) used CAP-TV, combined with a coaching session, to teach high school teachers four evidence-based classroom management practices. Teachers who learned using CAP-TV and received coaching implemented significantly more practices than colleagues who received a traditional in-person PD session as

measured by a researcher-created observation instrument. This successful, preliminary test of CAP-TV plus coaching provides a new angle when considering the effectiveness and applicability of a cognitive apprenticeship model of PD for teachers using multimedia. The current study extends this first empirical test of the CAP-TV approach by adding even more specific scaffolds to support implementation of high-quality instruction.

Feature 3: Sample Instructional Materials

Curriculum-linked PD is more effective than workshops that focus on general pedagogical strategies (Cohen & Hill, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). This approach to PD is guided by the need to put authentic instructional materials into the hands of teachers so they can have opportunities to learn and practice using the exact tools they would use when teaching (Fishman & Krajcik, 2003). Hill, Rowan, and Ball (2005) proposed a PD approach that balanced teachers' knowledge of curriculum and strategies for implementation was the best choice for designing and implementing an effective PD session. Thus, in this study, participants received curriculum materials for the key vocabulary terms and concepts that embed the evidence-based practices learned within the CAP-TVs. We call these materials CAP-Teacher Slides (CAP-TS). By giving teachers access to approximately 100 CAP-TS (each individual slide show corresponds to one vocabulary term from the curriculum), and also the flexibility to amend those materials, we support their readiness to implement the practices with fidelity balanced against their knowledge of how to discuss key terms and concepts. Sample CAP-TS can be accessed at www.VocabSupport.com.

The CAP-TS, in combination with the CAP-TVs are examples of educative curriculum materials (Davis & Krajcik, 2005). Educative curriculum materials are instructional materials intended to promote teacher learning and implementation of curriculum. Teachers who learn using these types of instructional materials are in a position to make informed decisions and also have sufficient knowledge to make needed adjustments when teaching. When integrated into a PD process, educative curriculum materials can provide the knowledge, pedagogy, and dispositions needed to successfully implement the new intervention or program (Davis & Krajcik, 2005). Educative curriculum materials such as CAP-TS are also compatible within a model of cognitive apprenticeship as these materials provide learners an opportunity to learn and practice within the situated context of their existing workplace.

For this study, researchers created a CAP-TS for every key vocabulary term and concept from the participating teachers' curriculum (n = ~100), thereby providing a clear illustration of how the teacher could choose to teach students across a range of terms. The slides embed sample questions to ask students and have blank spaces for the teachers to insert their own examples, ideas, and practices. Teachers were instructed to use the CAP-TS as a starting point in their planning and implementation of high-quality instruction that could be linked to inquiry and other activities expected in science. These slides bring the instructional practices from the CAP-TVs to life and provide the teachers with a running start to create powerful learning experiences for students. More details are provided in the "Method" section.

Feature 4: Feedback and Coaching

Coaching for teachers, live and virtual, has been a major focus of PD researchers in recent years (Hill et al., 2013). In a literature review on single-case design studies of the efficacy of performance feedback on teacher performance, Fallon, Collier-Meek, Maggin, Sanetti, and Johnson (2015) identified and analyzed 47 studies involving 169 individual cases. Twenty-nine of these studies demonstrated strong to moderate evidence of positive effects of feedback on the quality of teacher delivery of interventions. There was wide variation in the frequency and immediacy with which the feedback was provided. Coaching and feedback to learners is a key component of a cognitive apprenticeship model in that learners who receive specific feedback on performance within their situated work are in strong position to make any needed improvements (Brown, Collins, & Duguid, 1989).

Coaching in this study. In accordance with positive findings about the impact of feedback and coaching on teacher performance, the final stage of the CAP-PD process is to provide specific feedback and coaching based on live observations. After being observed, teachers received a feedback/coaching email from a member of the research team. Feedback was delivered electronically using data outputs from the dependent measure created for this study, the Classroom Teaching (CT) scan. In addition, researchers referred teachers back to CAP-TVs and CAP-TS when improvement was needed. More information is provided in the "Method" section.

Purpose of the study. The purpose of this study was to examine the effects of a multiple-component, multimedia PD package consisting of CAP-TVs, sample instructional materials in the form of (CAP-TS), and feedback and coaching on the vocabulary instruction of inclusive middle school science teachers. Thus, the research questions that guided this study were as follows:

> **Research Question 1:** Does implementation of a multi-component, multimedia PD package (CAP-PD) affect critical aspects of vocabulary instruction (e.g., time spent explicitly teaching vocabulary, number of vocabulary practices used with fidelity).
> **Research Question 2:** Does the PD package have strong social validity value for teachers?

The study is a partial replication of the CAP-T and CAP-TV intervention used in previous studies to support teachers' knowledge and implementation of evidence-based practices. In addition, we seek to further explore the hypothesis that cognitive apprenticeship is an effective theoretical model for developing multimedia-driven PD for teachers.

Method

To investigate these questions, we implemented a multiple-baseline across participants design. We observed participants' vocabulary instruction before and after the intervention and analyzed differences between the two phases. Following data collection, we conducted follow-up observations to determine the extent to which the participants maintained any changes in their practice after the intervention was no longer provided by the research team.

Participants

Based on school district and building administrator recommendations, we recruited teachers who taught in inclusive science classrooms, which were defined as class periods in which students with individualized education plans (IEPs) were present, and a special education coteacher provided support. The students with IEPs (n = 26) were grouped into specific sections of each teacher's schedule such that each teacher had at least one period with students with IEPs. Each teacher also taught at least one section without students with IEPs. The students with IEPs included students with specific learning disabilities (51%), attention deficit/hyperactivity disorder (ADHD; 28%), speech/language impairments (10%), emotional/behavioral disorders (9%), and autism spectrum disorders (2%).

Teacher A was a Caucasian female in her 40s who was in her sixth year of teaching. We observed her first block class, a seventh-grade life science course taught with a special education coteacher; there were 20 students representing a range of academic achievement levels, including seven students who had IEPs. Teacher B was a Hispanic male in his 20s in his second year of teaching. He taught science and mathematics at the school. We observed his second block, which was a cotaught seventh-grade life science course. There were 28 students in the class, and the students included four English language learners and eight students who had IEPs. Teacher C was a Caucasian female in her 20s during her fourth year of teaching. We observed her fourth block, a cotaught eighth-grade physical science course. This class was made up of 30 students, including 11 students with IEPs.

Researchers conducted a side study within the larger project to explore differences between how teachers taught sections with and without students with IEPs. It was assumed there would be pedagogical differences in how they taught sections when students with IEPs were present. However, results demonstrated there were no pedagogical differences at baseline. In practice, this meant the science teacher handled all of the instruction, and the special educator was relegated to noninstructional activities. We therefore made the decision to focus PD efforts on the general education teachers. The special education co-teachers were assigned to a particular group of students rather than to a particular general educator. In addition, special educators were intended to serve mainly as resources and supports for the students with disabilities; there was no expectation that the two teachers would plan or instruct together. Therefore, although the special educators were aware of the study and consented to be observed, we restricted provision of the PD package to the three general educators who served as the primary instructional planners and leaders in these classes. All three general educators volunteered and received stipends and access to PD materials in return for their participation.

Setting

The middle school chosen for this study was one of five middle schools in the district. It generally reflected the demographics of the division as a whole but was slightly more diverse in terms of race and socioeconomic status. Approximately 66% of the students in this middle school were White, 13% African American, and 16% Hispanic. Nearly 37% of students qualified for free or reduced lunches, 16% were identified as students with disabilities, and 5% qualified as English language learners.

Materials

Observation protocol. We measured vocabulary practices using the Classroom Teaching (CT) Scan (Kennedy, Rodgers, Gressick, Romig, & Mathews, 2017), an observational tool developed by the research team designed to capture class time use and instructional practices. It is grounded in the theory that explicit, direct instruction is the most effective method of teaching for students with disabilities (Archer & Hughes, 2011; Vaughn, Gersten, & Chard, 2000), and it allows the observer to record all teaching methods, practices, indicators of fidelity-quality in real time, and also qualitative notes stamped with the time during the lesson when the note was written.

During an observation, the observer determines the broad category of instruction being used by the teacher (e.g., vocabulary, behavior management, observing/assessing students, class procedures, general content instruction). When the observer clicks on the broad category title, a secondary list of specific practices appears. For example, if the vocabulary category was selected, a list including student-friendly definitions, discussion, examples, asking students to define term, morphological strategies, and several others can be selected. Finally, if the selected practice is an evidence-based practice for vocabulary instruction, a third menu pops onto the screen containing specific quality features for that practice. As an example, for the practice of discussion, the quality

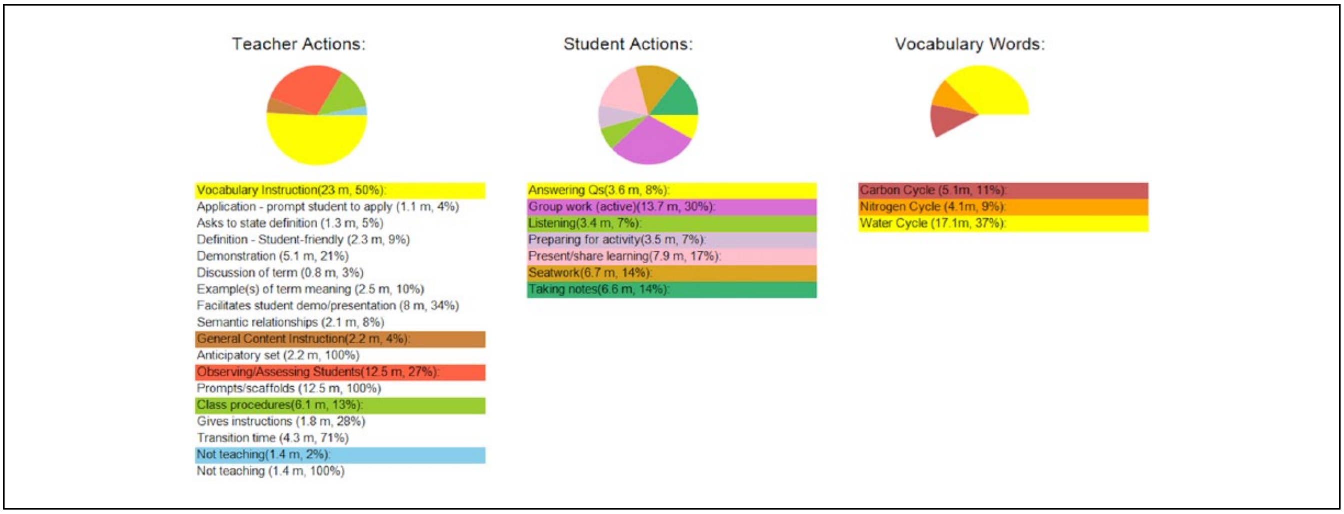

Figure 1. Pie chart output from one of Teacher A's observations.
Note. The left-hand column shows the instructional practices the teacher used in the lesson, the middle column lists the actions in which the students engaged, and the right-hand column shows the vocabulary words the teacher taught.

features are (a) uses appropriate, authentic context, (b) calls on students all around the room, (c) asks open-ended questions, (d) encourages student talk, and (e) incorporates student talk into the discussion (Ford-Connors & Paratore, 2015). All practices and quality features are identified using peer-reviewed literature, and defined explicitly in the observer codebook (contact the first author to obtain the codebook and to access the CT Scan).

The CT Scan records the amount of time the teacher uses the specific practice (with or without fidelity) synced to the vocabulary term or concept being taught. Prior to beginning the study, we conducted an extensive literature search and pilot observations to create a list of practices we expected middle school teachers to use in inclusive classrooms. For each practice that was research based, we constructed a list of three to five quality features based on best practices recommended in the literature.

After an observation, the CT Scan creates two visual outputs: a pie graph and timeline. The pie graph presents the percentage of time spent in six categories of teaching: General Content Instruction, Vocabulary Instruction, Observing/Assessing, Class Procedures, Behavior Management, and Not Teaching. It also shows the time students spent on various activities (e.g., taking notes or listening), the vocabulary words that were taught, which vocabulary practices were used, and counts of feedback statements and questions asked; Figure 1 shows a sample pie graph. The timeline is a visual representation of the class in real time. It shows what the teacher and students were doing, what vocabulary words were being taught, and an estimate of student engagement throughout the class; Figure 2 shows a sample timeline. The CT Scan is used as the measure of the main dependent variable in this study, but the data outputs are also part of the PD package delivered to participants in the form of coaching.

Intervention. We developed a multimedia PD package intended to improve teacher understanding of and ability to implement evidence-based vocabulary practices for students with disabilities. The PD package consisted of three components: (a) short multimedia instructional vignettes called CAP-TV (Rodgers et al., 2017), (b) sample instructional materials in the form of teacher slides for use during vocabulary instruction (CAP-TS), and (c) feedback and coaching using CT Scan outputs and feedback emailed to participants.

CAP-TV. The first component of the package was CAP-TV. Depending on the PD needs, the CAP-TV can be adapted to meet the specifications of the grade level and/or content area. For this study, we created five CAP-TVs: student-friendly definitions, examples and nonexamples (i.e., things that can be used to distinguish between related ideas or terms), morphological approaches (i.e., analyzing words using knowledge of their roots and affixes), semantic relationships among terms, and having high-quality discussions. Each CAP-TV was reviewed by two experts in the field of vocabulary instruction at the secondary level prior to use in this study. These five practices are not an exhaustive list of evidence-based vocabulary practices, but they were selected given their strong record of success in improving vocabulary-related outcomes for students with disabilities (Bryant, Goodwin, Bryant, & Higgins, 2003; Jitendra et al., 2004). In addition, these five practices used in combination may support teachers' broad goal of increasing the amount of time students receive high-quality vocabulary instruction because they encourage discussion of terms and using terms in and out of context and spanning different instructional contexts (Graves, 2004).

Based on findings from a pilot study, we hypothesized watching the CAP-TV alone may not be sufficiently powerful

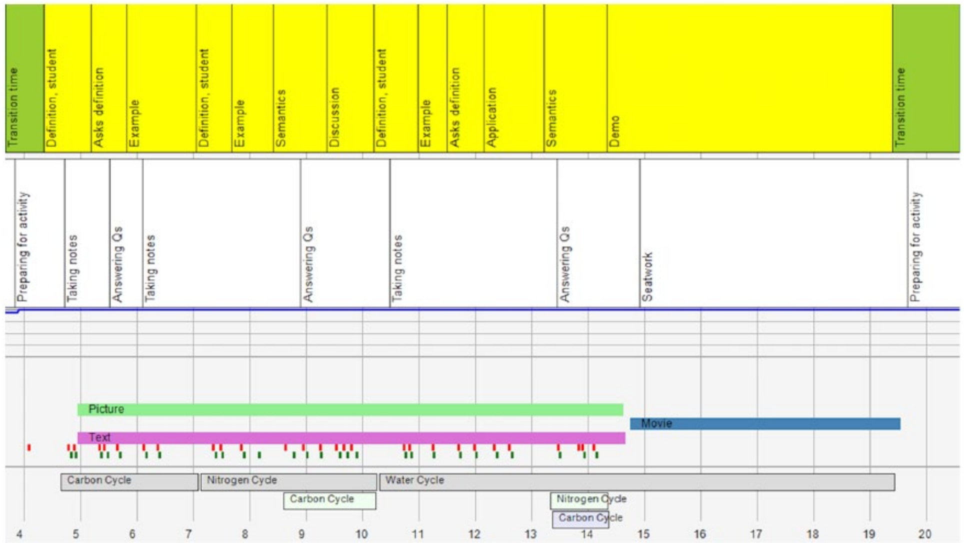

Figure 2. Timeline output from 16 min of the same observation.
Note. The top row shows teacher practices, the middle row shows student actions, and the bottom row shows opportunities to respond (OTRs), feedback statements, visual aids, and vocabulary words.

to result in sustained change in practice. Thus, we created a second arm to the CAP-PD intervention to help scaffold the teachers' implementation of the vocabulary practices in the specific context of the terms being taught.

Sample instructional materials. Based on the teachers' lists of vocabulary terms they planned to teach during the intervention period, we created CAP-TS that adhered to Mayer's (2009) cognitive theory of multimedia learning (CTML) principles and incorporated the strategies taught in the CAP-TV. Each slide in the CAP-TS contained sample speaker notes with a script for the teachers to consider, use, or adjust. The CAP-TS were uploaded to a shared dropbox folder. Teachers were instructed to use them as they saw fit, including making changes.

The first two elements of the PD package, watching CAP-TVs and receiving CAP-TS, are hypothesized to provide teachers with the knowledge and pedagogical examples and skill needed to provide high-quality vocabulary instruction embedded within their science teaching. However, we also hypothesized teachers would benefit from feedback on their performance and coaching to help sharpen fidelity of implementation.

Feedback and coaching. Following introduction of the CAP-TV and CAP-TS, we began sending feedback and coaching emails following each observation. Teachers received copies of the CT Scan pie graph and timeline outputs for that day, and in the body of the email, researchers provided feedback on strength areas and coaching in the form of specific suggestions for areas needing improvement. The structure of the email narratives was adapted from Hemmeter, Snyder, Kinder, and Artman (2011), but a conversational tone and personal connections were prioritized over structure. The main observer, one member of the research team, sent all of the feedback emails to ensure consistency across participants. Figure 3 contains a sample feedback email and response from a participant.

Dependent Variables

We collected data on two dependent variables: the time the teachers spent teaching vocabulary and the practices they used when teaching vocabulary. We also assessed social validity—the teachers' perspectives about the intervention practices. We describe these measures in the following paragraphs.

Time spent explicitly teaching vocabulary. Although there is no research-established amount of time teachers should spend teaching vocabulary, the consensus among researchers is that more exposure to words results in greater student learning (Elleman, Lindo, Morphy, & Compton, 2009; Ford-Connors & Paratore, 2015). Therefore, one of our key dependent variables related to teachers' instructional practices was the amount of time spent in explicit vocabulary instruction. For this measure, we counted any class time spent when the teacher was directly teaching vocabulary (e.g., giving notes, providing a demonstration) or when the teacher facilitated student demonstration of a vocabulary term. Because science is vocabulary rich, and terms and concepts are often seamlessly used during

Sample Coaching Email:

Dear (Participant Name),

As you know, the main point of this project is to help improve your vocabulary instruction. Today, vocab was 27% of the lesson, which is a solid number, and higher than what I have seen during your baseline observations. Within the domain of vocabulary, you used six different evidence-based practices for varying amounts of time – that is awesome! The most vocabulary practices you've ever used in a single day was two, so this is clear improvement. I am very pleased and proud of you. An area of improvement in the domain of vocab instruction is to make sure you follow up each individual practice by confirming students have learned. The easiest way to do this is to ask several students a question – I recommend spreading those questions around to target students, and those who never raise their hands or participate. You seem to have a group of students you call on regularly, but there are others I've never seen you interact with (that doesn't mean you haven't, I just haven't seen it).

Along with vocabulary, another data point we capture is the number of questions you ask, how many feedback statements you make to students, and how many questions are generated by your students. Today, you provided students with 30 opportunities to respond (OTRs), with all but one being a rote question (a question with a simple answer). Give some thought to how you can ask more thought provoking/challenging questions once in a while. It's very typical for the rote questions to outnumber the deeper ones, but it's nice to see a balance. 30 OTRs in 50 total minutes is an admirable number, but you can definitely do more. One thing to note is the 30 OTRs are entirely contained within the first 20 minutes of the lesson. This isn't surprising given that after the vocabulary part you transitioned into the independent work time. So really you did 30 OTRs in 20 minutes, which is impressive. To go along with the questions, you provided 17 generic feedback statements (e.g., Good, good job, or repeat their answer). A better option for feedback is to say things like, "Molly, yes, you nailed it, way to use your scientific reasoning." You don't have to do this every time, but once in a while will really help your students know why you are happy with them. Finally, 5 of your students asked a total of 7 questions, which is an area to improve on. You have 25 students, so when only 5 ask questions, it leaves me wondering the extent to which everybody is totally engaged. Based on what I can see most of your students are indeed engaged (see blue line on 2nd data output below), but it would be nice to see them participating a bit more. Fostering discussions about terms and concepts you are teaching will really encourage more participation. Will take some time to get students into a new routine. Spending more time engaged in vocabulary instruction or demonstrations of concepts is one way to allow time for this to develop.

In short, today's lesson looked a lot like what you normally do, with the exception of the vastly improved vocabulary piece. You have a routine of spending half or more of each day with students working on some sort of independent assignment/activity/project. I coded your time today as "proctoring" and "prompting/scaffolding." These aren't really instructional practices, so I don't have any feedback on this other than to shine a little light on the amount of time you're spending not directly teaching. There's nothing wrong with providing students with independent work time, but in my view, there's more value in you providing students with direct instruction and teacher directed practice opportunities. An idea when students are working in groups or even alone is to walk around and ask specific questions about the concepts they are working on – so you initiate discussions about the vocab/concepts with small groups or individuals. This "counts" as vocab instruction, and can be powerful for students. Plus it's a way for you to confirm their learning. These are just ideas.

You are an excellent teacher, and very positive role model for your students – especially the ladies to pursue careers in science. Keep up the good work.

Have a great day,
(Researcher Name)

Reply from Teacher:

(Researcher Name),

Thank you SO much for the feedback! :D I've never received detailed feedback like this before. I have noticed improvement in all of my classes! 5th period is definitely a tough one. But even the students that are struggling seemed more in tuned today, and the positive praise I could tell meant a ton to them! They really like the new slides I am using with the clear images.

I can tell the students are learning more than usual, and retaining what the terms mean. Some students who never participate are starting to open up. I'm also starting to see higher grades on quizzes and tests. I will continue to work on asking more questions and trying to get everybody involved.

THANK YOU!!! This is wonderful!

Sincerely,

(Teacher Name)

Figure 3. Sample feedback email and reply from teacher.

discussions and inquiry activities, our research team erred on the side of labeling instruction as vocabulary if a specific term or concept was being taught.

Vocabulary practices used with fidelity. We measured teachers' fidelity in using the vocabulary practices as a dependent variable. We tracked how many practices the teachers used per

class and how many of these practices were used with fidelity (i.e., all quality indicators were selected for that practice during the observation). Any practice a teacher used with fidelity at least once during the class was counted toward this measure. In other words, our measure potentially overcounted vocabulary instruction, but maintained a very high bar to be scored as vocabulary instruction with fidelity. Given the critical need for students to have multiple exposures to vocabulary terms in various contexts, we argue this decision is justified.

Interobserver agreement (IOA). To calculate IOA, two observers coded 25% of the observation sessions. This is more than the 20% of double coded observations recommended by Kratochwill and colleagues (2010). Agreement on percentage of class time spent on vocabulary instruction was defined as a result within 5%. Agreement on this dependent variable was 87%. For number of vocabulary practices taught with fidelity, exact agreement was defined as observers agreeing exactly on the number of practices, and adjacent agreement was defined as observers being no more than one practice apart. Exact agreement on this dependent variable was 60%, and adjacent agreement was 85%. We examined these data more closely, and we found that almost all disagreements were due to different responses on only one quality indicator. There was only one exception, in which the observers differed by two indicators.

Social validity. We assessed the social validity of the PD package through use of a researcher-developed survey the teachers completed after the intervention phase of the study was completed. Questions assessed teachers' perceptions of the usefulness of the PD package elements, changes in their practice associated with the PD, and general opinions about the PD as a whole. The survey consisted of 17 statements, and teachers indicated the extent to which they agreed with each on a scale of 1 (*strongly disagree*) to 5 (*strongly agree*). One statement ("I felt overwhelmed by the amount of new information") was reverse-coded. The survey also included seven open-ended questions to allow teachers to provide qualitative feedback.

Treatment Integrity

To measure treatment integrity, we tracked teachers' viewings of the CAP-TVs using www.EdPuzzle.com. Ed Puzzle is a free website where content can be uploaded, and viewers log in to access the material; based on login information, we determined that all teachers watched all five videos at least once. Also, during observations researchers noted whether the teachers were using the CAP-TS. All three teachers used the CAP-TS on days when they taught vocabulary. Finally, we asked teachers to send a reply once they received the feedback emails. All teachers replied to each feedback email and noted a general sense of thanks and appreciation for the positively worded notes.

Research Design

We used a multiple-baseline across participants design to assess the effectiveness of CAP-PD. This design uses participants as their own controls and provides evidence for the effectiveness of an intervention through repeated measurement of individuals' actions over time. In multiple-baseline studies, the intervention is introduced at staggered points in time across participants, thus allowing for demonstration of a functional relationship between implementation of the intervention and observed changes in the dependent variables (Horner & Odom, 2014).

We conducted observations for approximately 6 weeks, resulting in 29 observations for Teacher A, 30 for Teacher B, and 26 for Teacher C. We collected data under baseline conditions for 6, 10, and 15 days for Teachers B, A, and C, respectively. We collected data under intervention conditions for 13, 8, and 7 days for Teachers B, A, and C, respectively. We collected short-term maintenance data for 3 days for both Teachers B and A; 5 months later, we collected long-term maintenance data for 2 days for all teachers.

Baseline phase. Baseline data were collected through observations of the three teachers as described above. We told teachers we were studying the use of vocabulary instruction in middle school science classrooms but asked them to use their typical instruction. For the first 2 weeks, we observed multiple sections taught by each teacher to ensure the classes chosen as the focus for the remainder of the observations were representative of the teacher's typical instruction. An added benefit to the extra observations was that they decreased the amount of time required for the teachers to become accustomed to observers in their classrooms.

Intervention phase. After each observation, we examined visual graphs of the data. In accordance with guidelines on multiple-baseline studies from the What Works Clearinghouse (Kratochwill et al., 2010), we waited a minimum of five observations or until predictability was established (whichever was longer) for all teachers before introducing the intervention to any one teacher. To determine the order in which participants would receive the intervention, we used a random number generator. This randomization of participants to start points increases the study's internal and external validity which in turn strengthens the validity of the replication effects (Kratochwill & Levin, 2014). The participants received the intervention in this order: Teacher B, Teacher A, Teacher C. At the conclusion of the intervention phase, we ended all PD support—the teachers did not receive any more CAP-TS or feedback emails.

At the start of each intervention phase, we gave the teacher access to the CAP-TVs and the CAP-TS. We asked them to watch the CAP-TVs at least once but encouraged them to watch the videos as many times as necessary. We asked them to review the CAP-TS and revise them as needed to fit their

classrooms. Once the teachers had access to the PD materials, we began sending the feedback and coaching emails after the next observation date. Data from Ed Puzzle revealed the following numbers of CAP-TV views: Student-Friendly Definitions (Teacher A = 1, Teacher B = 2, Teacher C = 1), Examples and Nonexamples (A = 1, B = 2, C = 3), Morphological Approaches (A = 2, B = 3, C = 2), Semantic Approaches (A = 1, B = 3, C = 1), Discussion (A = 1, B = 1, C = 2).

Maintenance phase. We conducted short-term follow-up observations to measure maintenance of the target practices 2 weeks after the end of the final intervention phase for Teachers B and A. Five months after the short-term maintenance checks, we returned to the classrooms and conducted long-term maintenance observations for all three teachers. Data showed teachers did not access the CAP-TVs between the end of the first phase of the study and maintenance.

Data Analysis

To determine effectiveness of the PD package on the dependent variables, we employed visual inspection procedures to examine changes between the baseline and intervention phases in level, trend, and immediacy as defined by Kratochwill and Levin (2014). To calculate an estimate of the effect size of the intervention, we calculated the Tau effect size for each dependent variable. We did not need to calculate Tau-U because none of the teachers' baseline observations demonstrated positive or negative trend; therefore, we did not require the correction offered by Tau-U (Parker, Vannest, Davis, & Sauber, 2010). To calculate Tau, we entered the pertinent data from the graphs into the Tau calculator available at www.singlecaseresearch.org and analyzed the results. We chose to use Tau because it addresses change in level and trend, is well suited for small data sets, and has been shown to have more power than most other single-case effect sizes (Parker et al., 2010). We averaged scores on the social validity survey, and all members of the research team reviewed the qualitative responses.

Results

Results showed salutary changes in the two primary dependent variables. In addition, teachers reported substantial satisfaction with the goals, procedures, and outcomes of the project. We describe these outcomes in the following sections.

Time Spent Explicitly Teaching Vocabulary

Graphs of the results for the first dependent variable, time spent teaching vocabulary, are presented in Figure 4. To account for different lengths of classes, percentage of time was considered the key variable. All three teachers spent at least 4 times longer explicitly teaching vocabulary in their classes after the intervention than they had before (see Table 1). Before the intervention, the teachers spent an average of 10% (the equivalent of approximately 5 min) of the class on vocabulary instruction, and after the intervention they spent 46%. All three maintained higher percentages of vocabulary instruction at maintenance than they had at baseline, and their average in maintenance was 40%. Note that the last two maintenance points were collected over 5 months after the final intervention phase observation, providing strong evidence that the teachers incorporated the new strategies into their typical instruction.

Visual analysis of the graph clearly shows a change in level immediately after the start of the intervention for this dependent measure across all teachers. There is very little overlap of data points between the phases. Trend lines generally stay the same with slight evidence of trend reversal for Teachers B and A. In addition, these two teachers generally maintained changes after the intervention was removed; Teacher C's instruction related to this variable reverted to preintervention levels during the maintenance phase. However, this is accounted for by the fact that the teacher did not plan to teach vocabulary every day, which is appropriate for a middle school science teacher.

Calculations of Tau and statistical comparisons between the phase data indicated there was no baseline trend in any of the three cases. There was a significant difference between data in the baseline phase and the intervention phase in each case (all p values $< .01$), but the difference between the intervention phases and maintenance phases was not significant for either of the first two teachers (both p values $> .10$). The combined Tau effect size across cases was 0.95, with a 95% confidence interval of 0.62 to 1.28 ($z = 5.601$, $p < .001$), indicating a strong effect.

Number of Vocabulary Practices With Fidelity

Graphed results of teacher performance on number of practices used with fidelity are shown in Figure 5. Before the intervention, the teachers typically did not use any explicit vocabulary practices with fidelity, with few exceptions. After the intervention, they used an average of four practices with fidelity per class (see Table 1). They continued to employ the target practices at maintenance, using an average of six practices with fidelity per class, albeit for less time overall.

Teachers also varied in the types of practices they used. Teacher A typically used discussions and formal definitions. Teacher B used the most practices per class during baseline, and his most commonly used practices were asking students to state the definition, providing a student-friendly definition, and engaging students in discussion. Teacher C used the widest variety of practices across her baseline observations, but the most frequent she used by far were providing formal and student-friendly definitions.

Visual analysis of this graph shows a similar pattern as the one described previously. There is clear evidence of a change

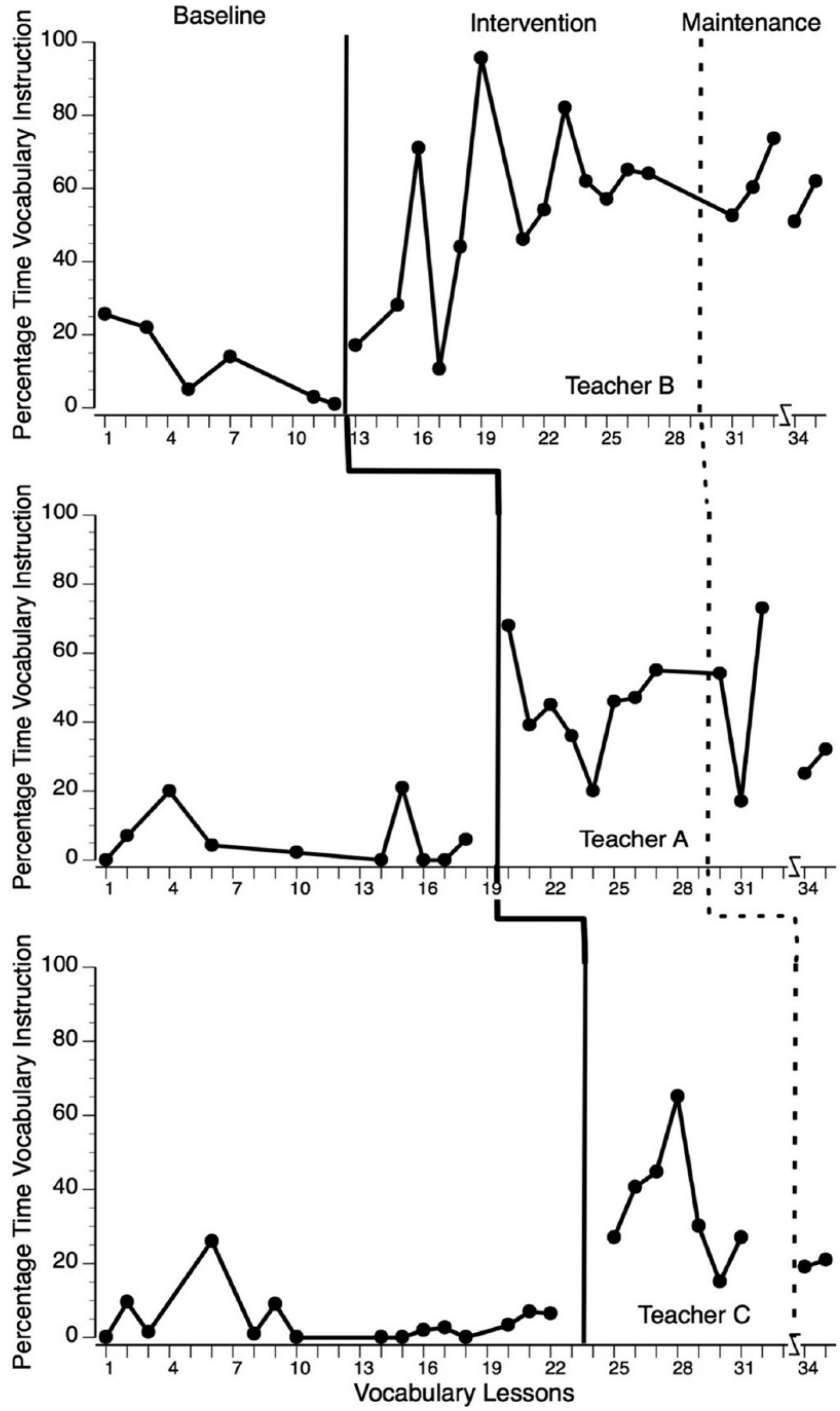

Figure 4. Graphs showing the percentage of time per class spent in explicit vocabulary instruction.

in level immediately upon start of the intervention across all teachers with no overlap of data points between the phases. Trend lines were generally flat, indicating consistency in teacher use of the practices both before and after the intervention. Changes in teacher instruction were largely maintained after the intervention was ended and continued to be

Table 1. Mean Values for Individual Teachers Results on Dependent Variable Measures by Phase.

Teacher	% class time on vocabulary instruction			Average number of vocabulary practices used with fidelity per class		
	BL	Int	Maint	BL	Int	Maint
B	13	58	60	.6	7	7
A	10	45	40	.07	5	7
C	7	36	20	0	2	5

Note. BL = baseline phase; Int = intervention phase; Maint = maintenance phase.

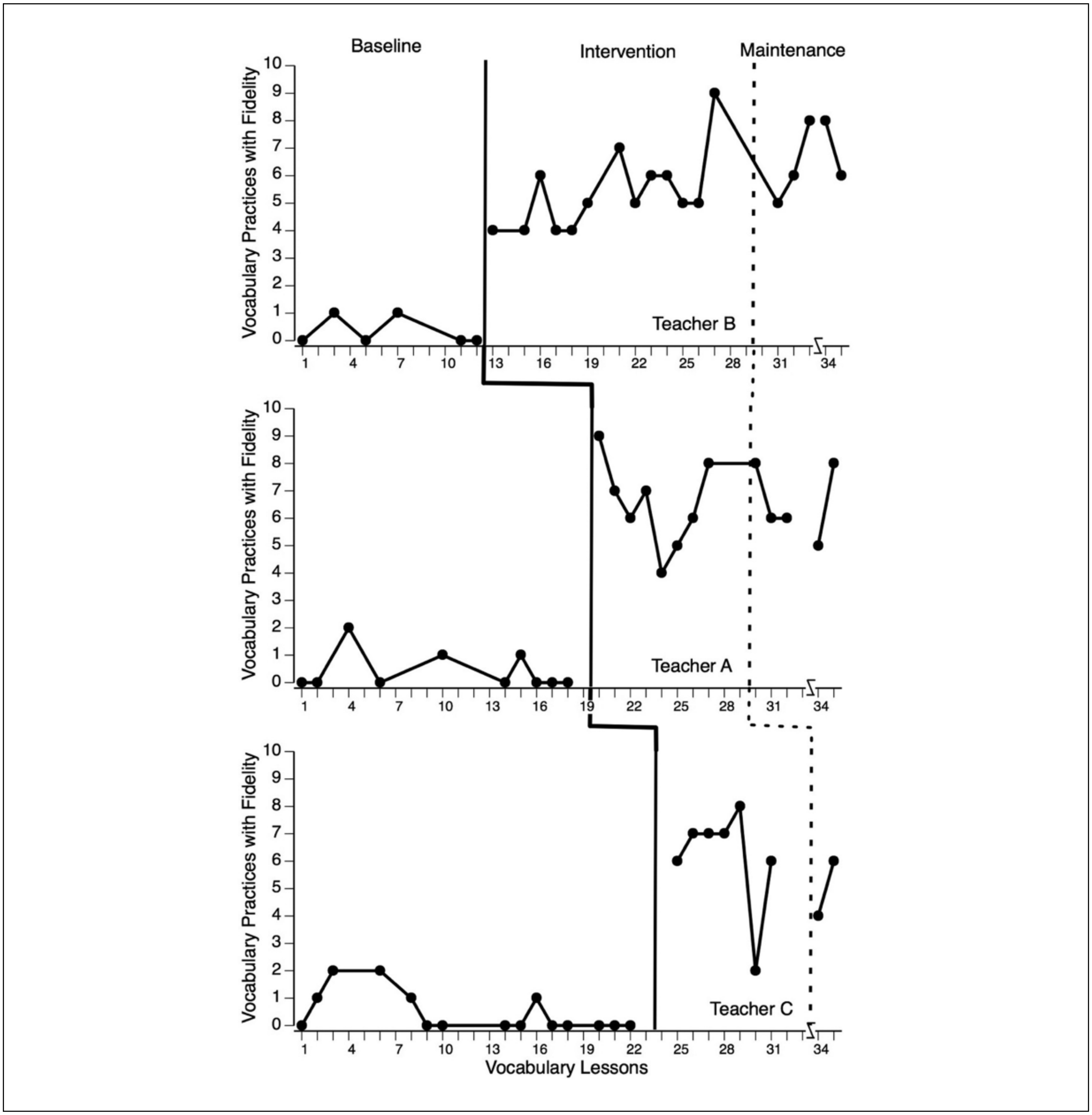

Figure 5. Graphs showing the number of unique vocabulary practices used with fidelity per class.

maintained in long-term maintenance observations conducted more than 5 months after the study ended.

Calculations of Tau and statistical comparisons between the phase data indicated again there was no baseline trend in any of the three cases. There was a significant difference between data in the baseline phase and the intervention phase in each case (all p values $\leq .01$), but the difference between the intervention phases and maintenance phases was not significant for either of the first two teachers (both p values $> .10$). The combined Tau effect size for number of practices with fidelity across cases was 0.87, with a 95% confidence interval of 0.54 to 1.21 ($z = 5.166$, $p < .001$), indicating a strong effect for this dependent variable as well.

Social Validity

Teachers were overwhelmingly positive about the intervention. The average agreement for all statements but one fell between 4 and 5, indicating that the teachers agreed or strongly agreed with all of those statements about the intervention. The only item that did not fall between 4 and 5 was the one that was reverse-coded; it assessed the extent to which teachers felt overwhelmed by the amount of information. On that statement, two of the teachers chose 1 (*strongly disagree*), and one chose 5 (*strongly agree*).

Results of the social validity survey were encouraging. All three teachers reported they learned a great deal from the intervention, they felt confident about creating their own CAP-TS for future units, and they felt they understood how to teach vocabulary explicitly and implement the morphological strategy of vocabulary instruction. Perhaps most importantly, all teachers agreed that this project helped their students learn and improved not only their vocabulary instruction but also their teaching more generally.

In their open-ended responses, two teachers wrote they appreciated the modeling videos that were embedded in the CAP-TV because they showed how to use the practices, as one teacher wrote, “in a ‘real-life’ scenario.” When asked about the feedback and coaching emails, all three teachers responded they appreciated the level of detail provided, and one teacher noted that the information was very “relevant” to her teaching. The teachers remarked on different aspects of the CAP-TS. One teacher wrote that they were “easy to edit and incorporate” in typical teaching; another liked that they were “tied to virtually all vocabulary teaching methods”; and the third noted the extensive review of the terms. When asked for additional comments, one teacher expressed concern about giving so much class time to explicit vocabulary instruction, writing that it took the place of some hands-on and inquiry-based activities and may not have allowed for as deep a study of the general concepts. Another expressed interest in seeing how the results of this instruction would compare across instructional units and time.

Discussion

First Step Toward a Larger Goal

In secondary science, as content and vocabulary become increasingly complex, students with disabilities tend to fall further behind their peers (Scruggs et al., 2010). In many ways, this gap can be attributed to deficits in the underlying skills that are essential for reading and comprehending complex content (Kennedy, Driver, Pullen, Ely, & Cole, 2013). That said, science teachers are taught and expected to provide inquiry-based instruction that aligns with state standards (which could be the Common Core State Standards) and/or the Next Generation Science Standards (NGSS; Quinn, Schweingruber, & Keller, 2012). Because the Common Core and NGSS are concept rich, teachers often find themselves stretched thin when trying to cover all of the content (Kesidou & Roseman, 2002) and can resort to less effective means of instruction (e.g., lecture) in an attempt to comply with pacing guides and other standards-driven mandates. Baseline data from teachers in this study provide preliminary evidence of this inefficient type of instruction.

Inquiry-driven instruction is an appropriate pathway to prepare students to be college and career ready (Quinn et al., 2012). However, students are unlikely to succeed in science courses if they do not understand the underlying concepts that facilitate meaningful participation in inquiry and other activities (Therrien et al., 2011). Lectures and having students read chapters from the textbook are examples of two oft-used methods teachers use to “ensure” students learn key vocabulary terms and concepts. Aforementioned data from the NAEP and Programme for International Student Assessment (PISA) are but two data points that demonstrate what many teachers are doing is not working. Therefore, high-quality vocabulary instruction is a prerequisite for supporting all students preparing to engage in inquiry-based activities.

An interesting question raised by our findings is whether science teachers spending up to half or more of a class on vocabulary instruction is a good thing. Although we succeeded at improving the quantity and quality of vocabulary instruction, this is not the final goal of science instruction. Instead, teachers should aim to incorporate the types of activities taught within this study into inquiry-based or hands-on instruction so they happen in a seamless manner. At baseline, teachers in this study were not spending much time (~10%) directly teaching vocabulary using quality practices, nor were they using any approaches that could be considered inquiry. Therefore, although most experts in science instruction would likely balk at a teacher spending 50% of a class period on vocabulary, we hypothesize this outcome is better than students spending the vast majority of time working independently on low-level activities like completing worksheets and copying definitions from the textbook glossary onto a study guide. In sum, this study is a first step toward a larger goal, in that we demonstrated science teachers can quickly learn to

implement higher quality vocabulary practices although abandoning ineffective practices. This is critical, as all teachers need to utilize the practices noted in this study, regardless of whether they teach students with or without disabilities.

Although it is difficult to make a broad statement about how middle school science teachers are collaborating (or not) with special education co-teachers based on three participants from this study, the finding at baseline that instruction did not differ for when students with IEPs were or were not present is troubling. As the teachers' instruction improved following access to the CAP-PD materials, their instruction improved for all of their students—not only those with IEPs. Although it would be better if teachers were providing individualized instruction as called for by respective students' IEPs, we were successful in at least helping them provide students with a higher quality of instruction than they were previously exposed to. This is an important, result to be explored with further research.

Expansion of Prior Research

This study is an extension and replication of prior work done on CAPs completed by Kennedy and his colleagues. In early studies (n = 13; for example, Kennedy, Hart, & Kellems, 2011), researchers tested the capacity of CAP-Ts to support only declarative knowledge of learners. As the line of research became more advanced, researchers used CAP-Ts to support learners' declarative and conditional knowledge of specific practices, such as curriculum-based measurement (Kennedy, Wagner, et al., 2016). The theory tested in each study was Mayer's (2009) cognitive theory of multimedia learning, and studies were largely completed in lab-type settings. In more recent work, researchers paired CAP-Ts with modeling videos to support declarative, conditional, and procedural knowledge (e.g., Kennedy, Hirsch, et al., 2017). The CAP-TV studies leveraged Mayer's (2009) theory in addition to a cognitive apprenticeship model (Brown et al., 1989).

The CAP-PD model is an example of how cognitive apprenticeship can be utilized in schools using multimedia. Teachers watch the CAP-TVs to both learn important information regarding new practices and also see those practices being implemented with fidelity. The videos feature real teachers, students, and learning situations to address the need for instruction to be situated. The teachers then receive highly specific feedback on their performance. By using Mayer's applied theory, we further bolster the opportunity for learners' success by presenting content in a learner-friendly format. In sum, we hypothesize that CAP-PD is an approach to improving teachers' implementation of evidence-based practices, consistent with cognitive apprenticeship.

Results of the social validity survey were encouraging. We believe this is largely due to the individualization of the PD materials for the teachers' specific content and the extensive interactions and specific feedback we provided. In addition, we purposefully limited the amount of time required for the teachers to participate in this PD; the instructional CAP-TVs were kept relatively short, and the CAP-TS we provided only required small adjustments before they were ready to be used in classes. The bulk of the PD consisted of time to practice using the interventions, which took place during normal instructional time and the amount of time it took teachers to read and reflect on the feedback emails and CT Scan outputs.

One concern we had at the beginning of this study was that the teachers might use the CAP-TS as a packaged curriculum and would not internalize the instructional practices taught in the CAP-TV. However, there were two indications this was not the case. One was that all three teachers almost immediately began personalizing the CAP-TS and even added to them, using information specific to their classes. This included making changes to certain examples proposed within the CAP-TS, changing images, and combining the study slides with slides the teachers had already created for the various terms. Second, the changes in teacher practice were maintained in most cases even after the teachers stopped teaching the content for which they had been created. This provides evidence for optimism that once teachers learn these strategies, they will be able to implement them successfully on their own after a period of scaffolded support. This is evidence that the CAP-TS are educational curriculum materials as described by Davis and Krajcik (2005).

Finally, our findings replicate those of previous classroom observation studies (McKenna et al., 2015; Swanson, 2008) that have shown teachers typically engage in very little explicit instruction and support the idea that general education teachers are not currently using instruction that is considered best practice. This is understandable given the lack of preparation teachers report in this area (Wei, Darling-Hammond, & Adamson, 2010). These observations provide additional evidence that in the contemporary climate of inclusion, more PD is needed for general educators to support use of high-quality practices.

Limitations

The results of this study are encouraging, but they should be considered along with its limitations. As with all single-case research, the small sample size limits the ability to generalize these findings to other populations. Replication is central to single-case research and, thus, necessary in this case. Also, we did not design this project in such a way that we could make causal statements about the impact of the PD on student outcomes. We hypothesize that as teachers use more evidence-based practices, their students' knowledge would increase, but we did not test that hypothesis in this study. Student outcomes should serve as a dependent variable in future research on this PD package.

One threat to the internal validity of this study is that the observers were all members of the research team, and none

of us were blind to the purpose or conditions of the study. We knew which phase each teacher was in during all observations. This imposes some doubt as to the validity of the findings. We sought to minimize this threat by developing a priori a codebook defining each instructional practice and its fidelity markers and by collecting second-observer data and comparing results. The magnitude of the differences in vocabulary instruction seen across teachers is strong evidence for the positive effects of this PD package although additional studies will be needed to verify the strength of those effects.

Another threat to internal validity in this study is that the main instrument used to measure teacher instruction, the CT Scan, has only been validity tested with the members of the research team, who are also its creators. It is unclear whether other observers would find the same results. Based on our interobserver reliability data, we are making changes to our codebook to define the quality indicators more specifically, and we are optimistic about our ability to train other observers to acceptable levels of reliability.

Finally, all observations were announced and scheduled with the participating teachers ahead of time. This was done to ensure we observed instructional days only, but it could potentially have resulted in teachers altering their normal instruction as a result of knowing they were going to be observed. However, the length of time and consistency of the baseline and intervention phase observations help control for this potential observer effect; the teachers knew they were going to be observed under both conditions, so that knowledge cannot account for the change in their performance. Also, even if the teachers were not using their normal instruction, this study shows that the teachers were capable of internalizing and implementing these practices at least well enough to use them during observations, which is still a promising finding.

Implications and Directions for Future Research

The successful implementation of CAP-PD has implications for how PD can be delivered to inservice teachers. The PD was differentiated based on content and prior knowledge, and it was delivered electronically so each teacher could engage with the content individually at his or her own chosen times. This method of creating and delivering PD is extremely flexible and could be easily adapted for different content areas and grades. In addition, once a CAP-TV is made, it can be used indefinitely with any teachers who need or want strengthening.

In future studies, we will consider making some changes to the CAP-PD system in response to teacher feedback and our own observations. One teacher requested the addition of an in-person meeting after receipt of the CAP-TV that could be used to clarify understandings and respond to questions. Another requested more detailed information about how to read and interpret the graphic outputs from the CT Scan. One of those teachers also felt unsure how to incorporate student accountability into the instructional practices in such a way that he or she would come out with a product, such as notes, he or she could use to review the information later. All of these suggestions and comments will be considered in future iterations of this PD package.

To strengthen findings, the package should be used by different teams in varied locations. In addition, future studies could be done in which observations are conducted by members of the team who are blind to the implementation dates of the intervention. Extension of the findings would be accomplished through replication of the study design with teachers in other content areas, in different grades, and with other types of evidence-based practices besides vocabulary (e.g., math, reading comprehension). Direct measures of teacher knowledge would also be a useful dependent measure. In this study, teachers' sustained use of the practices was a proxy for knowledge, but this should be directly measured.

Finally, the way in which the PD package was implemented for this study was quite time-consuming, requiring individualization of the CAP-TV and CAP-TS as well as numerous observations and feedback. Using the feedback template, sending emails did not take more than 5 to 10 min per observation. Future studies should be designed to identify the best combination of the elements of this PD, both in terms of which elements are included and what the dosage of those elements should be. The ideal combination would be one which maximizes effects on teacher and student performance at the same time as minimizing the workload of those providing the PD. Another vital line of research is whether and how CAP-PD affects student academic achievement on measures of vocabulary knowledge and science knowledge more generally.

Conclusion

We examined the impact of a multimedia, multicomponent PD package on the vocabulary instruction of three middle school science teachers using a multiple-baseline design. Results show the PD package had positive effects on the amount of time teachers spent on vocabulary instruction and the quality of their implementation of high-quality vocabulary instruction. This has implications for the delivery of PD to inservice teachers and the use of observations as coaching tools. PD packages like this one can help to bridge the gap between research and practice by training teachers on basic evidence-based practices for students with disabilities that can be used in inclusive classrooms.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This project was supported by an early career research and development grant made by the National Center for Special Education Research within the Institute of Education Sciences (#R324B130023) awarded to Michael Kennedy at the University of Virginia. No

specific endorsement of this project or its related products on the part of the U.S. Department of Education should be implied.

References

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, *61*, 315-343.

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction*. New York, NY: Guilford Press.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*, 498-520. doi:10.1037/0022-0663.93.3.498

Blanchard, M. R., LePrevost, C. E., Tolin, A. D., & Gutierrez, K. S. (2016). Investigating technology-enhanced teacher professional development in rural, high-poverty middle schools. *Educational Researcher*, *45*, 207-220. doi:10.3102/0013189x16644602

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3-15.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42.

Bryant, D. P., Goodwin, M., Bryant, B. R., & Higgins, K. (2003). Vocabulary instruction for students with learning disabilities: A review of the research. *Learning Disability Quarterly*, *26*, 117-128.

Carlisle, J. F., Kelcey, B., Rowan, B., & Phelps, G. (2011). Teachers' knowledge about early reading: Effects on students' gains in reading achievement. *Journal of Research on Educational Effectiveness*, *4*, 289-321. doi:10.1080/19345747.2010.539297

Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching*, *49*, 631-658.

Ciullo, S., Lembke, E. S., Carlisle, A., Thomas, C. N., Goodwin, M., & Judd, L. (2016). Implementation of evidence-based literacy practices in middle school response to intervention: An observation study. *Learning Disability Quarterly*, *39*(1), 44-57. doi:10.1177/0731948714566120

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(1), 3-14.

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, *60*, 8-19. doi:10.1177/0022487108327554

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181-199. doi:10.3102/0013189x09331140

Edmonds, M., & Briggs, K. L. (2003). The Instructional Content Emphasis Instrument: Observations of reading instruction. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31-52). Baltimore, MD: Brookes.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*, 1-44. doi:10.1080/19345740802539200

Ely, E., Kennedy, M. J., Pullen, P. C., Williams, M. C., & Hirsch, S. E. (2014). Improving instruction of future teachers: A multimedia approach that supports implementation of evidence-based practices. *Teaching and Teacher Education*, *44*, 35-43.

Ely, E., Pullen, P. C., Kennedy, M. J., & Williams, M. C. (2015). A multimedia tool to deliver professional development of vocabulary instruction. *Journal of Special Education Technology*, *30*(1), 59-72.

Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M. H., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, *81*, 227-246. doi:10.1177/0014402914551738

Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a computerized professional development program on teacher and student outcomes. *Journal of Teacher Education*, *61*, 302-312. doi:10.1177/002248711039556

Fishman, B. J., & Krajcik, J. S. (2003). What does it mean to create sustainable science curriculum innovations? *Science Education*, *87*, 564-573.

Fishman, B., Konstantopoulos, S., Kubitskey, B. W., Vath, R., Park, G., Johnson, H., & Edelson, D. C. (2013). Comparing the impact of online and face-to-face professional development in the context of curriculum implementation. *Journal of Teacher Education, 64*, 426-438. doi: 10.1177/0022487113494413

Ford-Connors, E., & Paratore, J. R. (2015). Vocabulary instruction in fifth grade and beyond: Sources of word learning and productive contexts for development. *Review of Educational Research*, *85*, 50-91. doi:10.3102/0034654314540943

Gersten, R., Vaughn, S., Deshler, D., & Schiller, E. (1997). What we know about using research findings: Implications for improving special education practice. *Journal of Learning Disabilities*, *30*, 466-476.

Graves, M. F. (2004). Teaching prefixes: As good as it gets. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 81-99). New York, NY: Guilford Press.

Hemmeter, M. L., Snyder, P., Kinder, K., & Artman, K. (2011). Impact of performance feedback delivered via electronic mail on preschool teachers' use of descriptive praise. *Early Childhood Research Quarterly*, *26*, 96-109. doi:10.1016/j.ecresq.2010.05.004

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher*, *42*, 476-487. doi:10.3102/0013189X13512674

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371-406.

Horner, R. H., & Odom, S. I. (2014). Constructing single-case research designs: Logic and options. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 27-51). Washington, DC: American Psychological Association.

Jitendra, A. K., Edwards, L. L., Sacks, G., & Jacobson, L. A. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children*, *70*, 299-322. doi:10.1177/001440290407000303

Kennedy, M. J., Alves, K. D., Miciak, J., Romig, J., Mathews, H. M., & Thomas, C. N. (2016). Evaluating the relationship between naturalistic content acquisition podcast views and course performance. *Teacher Education & Special Education*, *39*, 293-307. doi:10.1177/0888406416659529

Kennedy, M. J., Alves, K. D., & Rodgers, W. J. (2015). Innovations in the delivery of content knowledge in special education teacher preparation. *Intervention in School and Clinic*, *51*(2), 73-81. doi:10.1177/1053451215579268

Kennedy, M. J., Driver, M. K., Pullen, P. C., Ely, E., & Cole, M. T. (2013). Improving teacher candidates' knowledge of phonological awareness: A multimedia approach. *Computers & Education*, *64*, 42-51. doi:10.1016.j.compedu.2013.01.010

Kennedy, M. J., Hart, J. E., & Kellems, R. O. (2011). Using enhanced podcasts to augment limited instructional time in teacher preparation. *Teacher Education and Special Education*, *34*, 87-105. doi:10.1177/0888406410376203

Kennedy, M. J., Hirsch, S. E., Rodgers, W. J., Bruce, A., & Lloyd, J. W. (2017). Supporting high school teachers' implementation of evidence-based classroom management practices. *Teaching and Teacher Education. 63*, 47-57. doi: 10.1016/j.tate.2016.12.009

Kennedy, M. J., Rodgers, W. J., Gressick, W. T., Romig, J. E., & Mathews, H. M. (2017). The Classroom Teaching (CT) Scan: A flexible observation tool for general and special education teaching. In D. Edyburn (Ed.), *Advances in Special Education Technology (ASET): App development in special education*. Orlando, FL: Knowledge By Design, Inc.

Kennedy, M. J., Wagner, D., Stegall, J., Lembke, E., Miciak, J., Alves, K. D., . . . Hirsch, S. E. (2016). Using Content Acquisition Podcasts to improve teacher candidate knowledge of curriculum-based measurement. *Exceptional Children*, *82*, 303-320. doi:10.1177/0014402915615885

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching*, *39*, 522-549. doi:10.1002/tea.10035

Klingner, J. K. (2004). The science of professional development. *Journal of Learning Disabilities*, *37*, 248-255.

Klingner, J. K., Urbach, J., Golos, D., Brownell, M., & Menon, S. (2010). Teaching reading in the 21st century: A glimpse at how special education teachers promote reading comprehension. *Learning Disability Quarterly*, *33*, 59-74.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://files.eric.ed.gov/fulltext/ED510743.pdf

Kratochwill, T. R., & Levin, J. R. (2014). Enhancing the scientific credibility of single-case intervention research. Randomization to the rescue. In T. R. Kratochwill & J. R. Levin (Eds.), *Single case intervention research: Methodological and statistical advances* (pp. 122-144). Washington, DC: American Psychological Association.

Marsh, B., & Mitchell, N. (2014). The role of video in teacher professional development. *Teacher Development*, *18*, 403-417. doi:10.1080/13664530.2014.938106

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

McKenna, J. W., Shin, M., & Ciullo, S. (2015). Evaluating reading and mathematics instruction for students with learning disabilities: A synthesis of observation research. *Learning Disability Quarterly*, *38*, 195-207. doi:10.1177/0731948714564576

Mutch-Jones, K., Puttick, G., & Minner, D. (2012). Lesson study for accessible science: Building expertise to improve practice in inclusive science classrooms. *Journal of Research and Science Teaching*, *48*, 1012-1034. doi:10.1002/tea.21034

National Center for Education Statistics. (2012). *The nation's report card: Vocabulary results from the 2009 and 2011 NAEP reading assessment* (NCES 2013 452). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2010). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy*, *42*, 284-299. doi:10.1016/j.beth.2010.08.006

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*, 921-958. doi:10.3102/0002831207308221

Quinn, H. R., Schweingruber, H., & Keller, T. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academy Press.

Reed, D. K., Petscher, Y., & Foorman, B. R. (2016). The contribution of vocabulary knowledge and spelling to the reading comprehension of students who are and are not English language learners. *Reading and Writing*, *29*, 633-657. doi:10.1007/s11145-015-9619-3

Rodgers, W. J., Kennedy, M. J., Alves, K. D., & Romig, J. E. (2017). A multimedia tool for teacher education and professional development. In C. L. Martin & D. Polly (Eds.), *Handbook of research on teacher education and professional development* (pp. 285-296). IGI Global. doi:10.4018/978-1-5225-1067-3.ch015

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator*, *36*(1), 12-20.

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, *37*, 525-559.

Scruggs, T. E., Mastropieri, M. A., Berkeley, S., & Graetz, J. E. (2010). Do special education interventions improve learning of secondary content? A meta-analysis. *Remedial and Special Education*, *31*, 437-449. doi:10.1177/0741932508327465

Swanson, E. A. (2008). Observing reading instruction for students with learning disabilities: A synthesis. *Learning Disability Quarterly*, *31*, 115-133.

Swanson, E. A., Solis, M., Ciullo, S., & McKenna, J. W. (2012). Special education teachers' perceptions and instructional practices in response to intervention implementation. *Learning Disability Quarterly*, *35*, 115-126. doi:10.1177/0731948711432510

Therrien, W. J., Taylor, J. C., Hosp, J. L., Kaldenberg, E. R., & Gorsh, J. (2011). Science instruction for students with learning disabilities: A meta-analysis. *Learning Disabilities Research & Practice*, *26*, 188-203.

Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children*, *67*, 99-114.

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. Dallas, TX: National Staff Development Council.

Yoon, K. S., Duncan, T., Lee, S., Scarloss, W.-Y., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No.033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Yore, L. D., Hand, B. M., & Florence, M. K. (2004). Scientists' views of science, models of writing, and science writing practices. *Journal of Research in Science Teaching*, *41*, 338-369. doi:10.1002/tea.20008

Zhang, M., Lundeberg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education*, *27*, 454-462. doi:10.1016/j.tate.2010.09.015

Author Biographies

Michael J. Kennedy is an Assistant Professor of Special Education at The University of Virginia's Curry School of Education. The core of his research is applications of multimedia to support teachers' learning and implementation of evidence-based practices.

Wendy J. Rodgers is a doctoral candidate at the University of Virginia. She studies classroom observation and effective practices in inclusive co-taught classrooms.

John Elwood Romig is a doctoral candidate at the University of Virginia. He studies teacher preparation, teacher observation, and writing instruction for students with disabilities.

John Wills Lloyd has taught and conducted research at the University of Virginia's Curry School of Education since 1978. He has focused on the learning and behavior problems experienced by students with disabilities and instructional practices that teachers can employ to help students overcome those issues.

Mary T. Brownell is a Professor of Special Education at the University of Florida and Director of the Collaboration for Effective Educator Development, Accountability and Reform (CEEDAR Center). Her research interests include teacher education, teacher assessment, and professional development, primarily in the area of reading.